

CLAIMS

- 1 1. A method for searching a corpus of documents,
2 comprising:
 - 3 defining a knowledge domain;
 - 4 identifying a set of reference documents in the
5 corpus pertinent to the domain;
 - 6 inputting a first query;
 - 7 searching the corpus using the set of reference
8 documents to find one or more of the documents in the
9 corpus that contain information in the domain relevant to
10 the first query; and
 - 11 adding at least one of the found documents to the
12 set of reference documents for use in searching the
13 corpus for information in the domain relevant to a
14 second, subsequent query.
- 1 2. A method according to claim 1, wherein inputting the
2 first query comprises inputting one or more search terms.
- 1 3. A method according to claim 2, wherein searching the
2 corpus comprises finding lexical characteristics of terms
3 in the reference documents and refining the search terms
4 using the lexical characteristics.
- 1 4. A method according to claim 1, wherein inputting the
2 first query comprises specifying one or more documents
3 representative of the information to be found in the
4 corpus.
- 1 5. A method according to claim 1, wherein searching the
2 corpus comprises searching the corpus to find the
3 documents that contain the information relevant to the
4 query and ranking the found documents by comparing them
5 to the set of reference documents.

1 6. A method according to claim 5, wherein ranking the
2 found documents comprises evaluating a textual
3 resemblance between the found documents and the reference
4 documents.

1 7. A method according to claim 5, wherein ranking the
2 found documents comprises assessing links between the
3 found documents and the reference documents.

1 8. A method according to claim 5, wherein adding the at
2 least one of the found documents comprises adding at
3 least the document having the highest ranking.

1 9. A method according to claim 1, wherein adding the at
2 least one of the found documents comprises removing one
3 of the documents from the set responsive to adding the at
4 least one of the found documents.

1 10. A method according to claim 9, and comprising
2 tracking a level of relevance of the reference documents
3 to the queries, and wherein removing the one of the
4 documents comprises removing one of the reference
5 documents whose tracked level of relevance is low.

1 11. A method according to claim 1, wherein the corpus
2 comprises at least a part of the World Wide Web, and the
3 documents comprise Web pages, and wherein searching the
4 corpus comprises conveying the query to one or more Web
5 search engines.

1 12. A method according to claim 11, wherein inputting
2 the first query comprises receiving the query from a user
3 of a pervasive device, and wherein searching the corpus
4 comprises searching while the device is disconnected from
5 the Web.

1 13. A method according to claim 1, wherein identifying
2 the set of reference documents comprises opening one or
3 more files of a knowledge base on a computer in which
4 data regarding the reference documents are saved.

1 14. A method according to claim 13, wherein identifying
2 the set of reference documents comprises identifying the
3 set of documents used by a first user in searching the
4 corpus, and wherein opening the one or more files
5 comprises copying the files for use by a second user in
6 searching the corpus for information in the domain.

1 15. A method for searching a corpus of documents
2 containing terms, comprising:

3 defining a knowledge domain;
4 identifying a set of reference documents in the
5 corpus pertinent to the domain;
6 finding lexical characteristics of the terms in the
7 reference documents;

8 inputting a search query;
9 refining the search query using the lexical
10 characteristics; and
11 searching the corpus to find information in the
12 domain responsive to the refined query.

1 16. A method according to claim 15, wherein finding the
2 lexical characteristics comprises finding lexical
3 affinities among the terms.

1 17. A method according to claim 16, wherein the search
2 query comprises search terms, and wherein refining the
3 search query comprises adding to the search terms further
4 terms found to have lexical affinity to the search terms.

- 1 18. A method for searching a corpus of linked documents
- 2 containing terms, comprising:
 - 3 defining a knowledge domain;
 - 4 identifying a set of reference documents in the
 - 5 corpus pertinent to the domain;
 - 6 inputting a search query;
 - 7 searching the corpus to find one or more of the
 - 8 documents in the corpus that contain information relevant
 - 9 to the query;
 - 10 evaluating a textual resemblance between the found
 - 11 documents and the reference documents so as to assign
 - 12 respective textual scores to the found documents;
 - 13 assessing links between the found documents and the
 - 14 reference documents so as to assign respective
 - 15 topological scores to the found documents; and
 - 16 ranking the found documents with respect to their
 - 17 relevance to the domain responsive to the textual scores
 - 18 and the topological scores.
- 1 19. A method according to claim 18, wherein evaluating
- 2 the textual resemblance comprises assessing, for each of
- 3 a plurality of the terms in the found documents, a
- 4 respective frequency of occurrence in the reference
- 5 documents.
- 1 20. A method according to claim 18, wherein the
- 2 documents comprise World Wide Web pages, and wherein
- 3 assessing the links comprises generating a graph of the
- 4 links between the pages and calculating authority weights
- 5 of the nodes of the graph.
- 1 21. Apparatus for searching a corpus of documents,
- 2 comprising:

3 a memory, adapted to store an identification of a
4 set of reference documents in the corpus pertinent to a
5 predefined knowledge domain; and

6 a search processor, which responsive to receiving a
7 first query as input, is adapted to search the corpus
8 using the set of reference documents to find one or more
9 of the documents in the corpus that contain information
10 in the domain relevant to the first query, and to add at
11 least one of the found documents to the set of reference
12 documents stored in the memory for use in searching the
13 corpus for information in the domain relevant to a
14 second, subsequent query.

1 22. Apparatus according to claim 21, wherein the
2 processor is adapted to find lexical characteristics of
3 the terms in the reference documents and to refine the
4 search query using the lexical characteristics.

1 23. Apparatus according to claim 21, wherein the
2 processor is adapted to receive the documents found to
3 contain the information relevant to the query and to rank
4 the found documents by comparing them to the set of
5 reference documents.

1 24. Apparatus according to claim 23, wherein the
2 processor is adapted to add to the corpus at least the
3 document having the highest ranking.

1 25. Apparatus according to claim 21, wherein the
2 processor is adapted to remove one of the documents from
3 the set responsive to adding the at least one of the
4 found documents.

1 26. Apparatus according to claim 21, wherein the corpus
2 comprises at least a part of the World Wide Web, and the

3 documents comprise Web pages, and wherein the processor
4 is adapted to search the corpus by conveying the query to
5 one or more Web search engines.

1 27. Apparatus according to claim 21, wherein the
2 processor is adapted to receive the query over a
3 communication link from a user of a pervasive device, and
4 to search the corpus while the communication link is
5 disconnected.

1 28. Apparatus for searching a corpus of documents
2 containing terms, comprising:

3 a memory, adapted to store an identification of a
4 set of reference documents in the corpus pertinent to a
5 predefined knowledge domain; and

6 a search processor, which is adapted to find lexical
7 characteristics of the terms in the reference documents,
8 and responsive to receiving a query as input, is adapted
9 to refine the search query using the lexical
10 characteristics and to search the corpus to find
11 information in the domain responsive to the refined
12 query.

1 29. Apparatus for searching a corpus of linked documents
2 containing terms, comprising:

3 a memory, adapted to store an identification of a
4 set of reference documents in the corpus pertinent to a
5 predefined knowledge domain; and

6 a search processor, which responsive to receiving a
7 query as input, is adapted to search the corpus to find
8 one or more of the documents in the corpus that contain
9 information relevant to the query, to evaluate a textual
10 resemblance between the found documents and the reference
11 documents so as to assign respective textual scores to

12 the found documents, to assess links between the found
13 documents and the reference documents so as to assign
14 respective topological* scores to the found documents, and
15 to rank the found documents with respect to their
16 relevance to the domain responsive to the textual scores
17 and the topological scores.

1 30. A computer software product for searching a corpus
2 of documents, the product comprising a computer-readable
3 medium in which program instructions are stored, which
4 instructions, when read by a computer, cause the computer
5 to receive a definition of a knowledge domain and an
6 identification of a set of reference documents in the
7 corpus pertinent to the domain, and further cause the
8 computer, responsive to a first query, to search the
9 corpus using the set of reference documents to find one
10 or more of the documents in the corpus that contain
11 information in the domain relevant to the first query,
12 and to add at least one of the found documents to the set
13 of reference documents for use in searching the corpus
14 for information in the domain relevant to a second,
15 subsequent query.

1 31. A product according to claim 30, wherein the corpus
2 comprises the World Wide Web, and the documents comprise
3 Web pages, and wherein the instructions cause the
4 computer to search the Web by conveying the query to one
5 or more Web search engines.

1 32. A product according to claim 31, wherein the
2 instructions cause the computer to receive the first
3 query from a pervasive device, and to search the Web
4 while the pervasive device is disconnected from the Web.

1 33. A computer software product for searching a corpus
2 of documents, the product comprising a computer-readable
3 medium in which program instructions are stored, which
4 instructions, when read by a computer, cause the computer
5 to receive a definition of a knowledge domain and an
6 identification of a set of reference documents in the
7 corpus pertinent to the domain and to find lexical
8 characteristics of the terms in the reference documents,
9 and further cause the computer, responsive to a query, to
10 refine the search query using the lexical characteristics
11 and to search the corpus to find information in the
12 domain responsive to the refined query.

1 34. A computer software product for searching a corpus
2 of documents, the product comprising a computer-readable
3 medium in which program instructions are stored, which
4 instructions, when read by a computer, cause the computer
5 to receive a definition of a knowledge domain and an
6 identification of a set of reference documents in the
7 corpus pertinent to the domain, and further cause the
8 computer, responsive to a query, to search the corpus to
9 find one or more of the documents in the corpus that
10 contain information relevant to the query, to evaluate a
11 textual resemblance between the found documents and the
12 reference documents to assign respective textual scores
13 to the found documents, to assess links between the found
14 documents and the reference documents to assign
15 respective topological scores to the found documents, and
16 to rank the found documents with respect to their
17 relevance to the domain responsive to the textual scores
18 and the topological scores.